DOCUMENT RESUME

ED 211 976 CS 206 690

AUTHOR Thomas, David; Donlan, Dan
TITLE Correlations between Holistic and Quantitative Methods of Evaluating Student Writing, Grades 4-12.
PUB DATE Mar 80
NOTE 31p.; Paper presented at the combined Annual Meeting of the Conference on English Education and the Secondary School English Conference (Omaha, NE, March 27-29, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Ability; Comparative Analysis; Correlation; Developmental Stages; Elementary Secondary Education; *Evaluation Methods; *Holistic Evaluation; *Student Development; Student Evaluation; *Writing Evaluation; *Writing Research; *Writing Skills

ABSTRACT

A random sample of 175 compositions on the same topic (a lost suitcase) was used in a study examining the correlations between holistic (single impression) and quantitative methods of evaluating student writing. The sample contained 25 papers from each of the following grade levels: four/five, and seven through twelve. A panel of three readers made the holistic evaluations of the papers during a regional school district writing assessment. Mean quality scores from this evaluation indicated a linear increase across grade levels in the quality of the compositions. The quality scores for the papers were then correlated with six extrinsic variables: (1) composition length (number of words), (2) number of T-units, (3) words per T-unit, (4) number of items in suitcase, (5) number of described items, and (6) number of nonclothing items. It was found that the number of words was the variable most highly correlated with quality despite grade level, and that the number of items described correlated the least with quality. Correlations showed that the six variables seemed to be independent of one another. The analyses also indicated a regression in writing abilities between the eighth and the tenth grades. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED211976

D133

CORRELATIONS BETWEEN HOLISTIC AND QUANTITATIVE METHODS OF EVALUATING STUDENT WRITING, GRADES 4-12

David Thomas

University of California, Riverside

Dan Donlan

University of California, Riverside

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David Thomas
Dan Donlan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS206690

CORRELATIONS BETWEEN HOLISTIC AND QUANTITATIVE METHODS OF EVALUATING STUDENT WRITING, GRADES 4-12

Background

A major breakthrough in the search for objective, sensitive and reliable instruments in the evaluation of written composition across the grade levels came in the 1960's with the development by two independent researchers, Walter Loban (1961) and Kellogg Hunt (1965), of a quantifiably significant syntactical unit. Both investigators observed that clause length within sentences tended to increase in student compositions in succeeding grade levels. Loban described this trend in terms of an index of elaboration based on what he chose to call a communication unit. He defined the communication unit as a basic unit of discourse, consisting of a single independent clause and all its attendent modifiers. Because he was working with both oral and written expression, he extended his definintion to include non-clausal responses to questions, responses for example, as "yes" or "no" to a question such as "Do you have paper and pen?"

Hunt, analyzing only written compositions, evolved what he called a terminal or T-unit, essentially the same as Loban's communication unit, though adhering strictly to the requirement of at least one subject and one predicate. The studies conducted by Loban (1961, 1976) and Hunt (1965, 1970) and other researchers such as O'Donnell, Griffin and Norris (1967) and O'Hare (1973) consistently demonstrated that

the mean length of T-units increased in student compositions grade level by grade level. The isolation of the T-unit and the observation that its length increases significantly with grade level set the stage for further research on the structure (and evolution) of the T-unit (in its expanding course) through the grade levels.

In constrast with this relatively quantifiable and objective syntactical analysis is a more impressionistic approach to the assessment of student writing. For state and national assessors dealing with hundreds and thousands of student essays, individual analysis by T-unit is impractical. Cooper and Lloyd-Jones in Cooper and Odell (1977) and Myers (1980) have each described what have proven to be valid and reasonably reliable methods for the holistic evaluation and ranking of compositions.

Holistic evaluation relies on trained readers to rank or sort samples of writing quickly and impressionistically according to a specific rubric or grading scale. The rubric is drawn up inductively from samples of writing in the mode for which the scale is being constructed and so is program specific (Cooper & Odell, 1977). Both Cooper and Lloyd-Jones assert a greater validity for whole composition assessment over the more quantifiable analytic techniques. Because so much of the reader's judgment impressionistic, though, inter-reader reliability in holistic scoring may be open to question. Lloyd-Jones, however, notes that reliability can be improved to an acceptable level when raters from like backgrounds are systematically trained. The studies which he cites show that training has brought inter-reader

reliability into the range between .73 and 1.0. Cooper (1977) also describes holistic scoring guides which he contends can achieve a scoring reliability as high as .90 for individual writers.

The Problem

Two such contrasting approaches to the assessment of written composition, i.e., T-unit vs. holistic analysis, might profitably bear comparison when used independently on the same body of student essays. Positive correlation would tend to reinforce our confidence both in the instruments as well as in the assessment of the scored sample. Just such an opportunity presented itself to us in the course of completing a Title IV-C funded, two county assessment of student growth in writing. Investigators were able to obtain samples of writing drawn from students in grades four through twelve from a variety of classes and schools which provided the basis for the present study. Of particular interest was not only the correlation between holistic and analytic scoring methods but also correlation by grade level of holistic ratings with six quantifiable syntactic and invention features.

Procedure

Each of fifty-two teachers who had participated in the Title IV-C writing project was asked to submit a writing sample from one class. The teachers were asked to exclude from consideration classes for special students, classes for example such as journalism, special education and honors. All students were to write on the same topic under scripted conditions (see Figure 1). The topic was designed

Insert Figure 1 about here

so that any student who had reached the acquisition level (independent composing) could write about it. The name, Terry Lee, was selected because it was androgynous and non-ethnic. The topic concerned a lost suitcase and the students were asked to write a letter to Ms. Martin regarding a lost suitcase. Students were allowed forty minutes to complete the assignment.

Twenty-nine teachers responded to the request for a writing sample. There were 853 papers from grades four through twelve. (No sixth grade papers were received.) Investigators took a random sample of 175 papers selected by grade level for the study. The sample distribution may be seen in Table 1. The writing was done under

Insert Table 1 about here

scripted conditions without any assistance from the teacher beyond clarification of the assignment. The papers were collected and holistically scored by three readers using reliable methods similar to those described by Miles Myers in his book A Procedure for Writing Assessment and Holistic Scoring (1980). After the papers were scored, investigators analyzed them in terms of six quantitative factors

SYNTACTICAL FEATURES

1. Number of words
2. Number of T-units
3. Number of words per T-unit

INVENTION INDICES

4. Number of items included in the suit case
5. Number of items in the suitcase which were described
6. Number of non-clothing items included in the suitcase

These features were necessarily extrinsic to the rubric and the concerns of the trained readers who holistically scored the essays.

Results

Tabulations were made on the overall quality score as well as on each of the six extrinsic variables, and mean scores for each variable per grade level were computed. With respect to number of words, data mean score increases with grade (Figure 2). Though the

---

Insert Figure 2 about here

---

dramatic increase in score between the eleventh and twelfth grades may not accurately reflect student writing insofar as several samples were submitted by students who had taken English honors as juniors and were demonstrably superior writers, the increase in holistic score is otherwise consistent with the increase in grade level, except for a slight plateau between the seventh and eighth grades.

The mean number of T-units per composition tended generally to increase across grade levels also, though a decline is noticeable between the eighth and tenth grades (see Figure 3). Similar trends may also be observed in Figure 4, which charts the mean number of

T-units by grade level. Words per T-unit, the most reliable syntactical measurement, (Figure 5) appears generally to increase,

---

Insert Figures 3, 4, and 5 about here

---

and following the trend of the preceding measures, shows a slump, both in the eighth and ninth grades. One may here begin to see a pattern in the syntactical features measured in the study, a pattern generally of a growth that is interrupted in the period between the seventh and tenth grades by regressions.

As an aspect of invention and imagination, investigators were interested in the number of items the writer would elect to include in the hypothetical suitcase. Again, the number of items increases across grade levels with a regression occurring in the seventh and ninth grades (see Figure 6).

Insert Figure 6 about here

The mean number of items that were described by the writer (rather than merely listed) demonstrates again a grade level growth, with the ninth grade in this instance incurring a particularly dramatic regression (see Figure 7).

Insert Figure 7 about here

The final measurement, the number of non-clothing items invented by the student to put into his lost suitcase (Figure 8), proved also

---

Insert Figure 8 about here

---

to be a factor correlating with increasing grade level despite slumps in the seventh and tenth grades.

Aside from the plateauing which we observed in the mean holistic scores in the seventh and eighth grades, quality scoring does not immediately reveal the regressions indicated by our quantitative measures. Investigators then decided to correlate (Pearson correlation coefficient) the holistic scores with each of the six variables to determine more precisely where quantitative measurements tended to confirm or contradict the quality scoring.

As can be seen in Table 2, the significance of the overall correlation or agreement between the holistic scores and each of the six variables is quite high. The highest correlation is with number of words; the lowest, with number of items described. However, as

---

Insert Table 2 about here

---

will be seen in Table 3, correlations at individual grade levels vary some, variables at times becoming negatively correlated with the holistic score. Generally, number of words is the best predictor of quality; number of items, the worst.

Insert Table 3 about here

In Figure 9 the correlations of the holistic scores with mean number of words, generally high, drops drastically in the ninth

Insert Figure 9 about here

grade and remains relatively low through the tenth. The correlations of holistic scores with the mean number of T-units, as shown in Figure 10, tends generally to increase to grade 8, drops sharply in grade 9,

Insert Figure 10 about here

and continues erratically through grade 12. The correlation of holistic scores with words per T-unit (Figure 11), is so erratic at

Insert Figure 11 about here

individual grade levels as to be negatively correlated with holistic scores with papers from eighth, tenth and twelfth grade students. In Figure 12 may be seen a general decline in the correlation of

Insert Figure 12 about here

holisitic scores with number of items by grade level, particularly accentuated in the ninth and tenth grades. A similar decline in correlation with items described by grade level may be observed in Figure 13 though here with important drops in the eighth and tenth

---

Insert Figure 13 about here

---

grades. In our final graph (Figure 14) correlation of holistic scores with number of non-clothing items shows a decline through the tenth grade, whereupon it rises significantly in the eleventh grade and again to a lesser degree in the twelfth.

Insert Figure 14 about here

One may gather from these correlation data that each of the features has its own particular profile. To bring these periods into greater focus, we isolated by grade level those correlations which were by a two-tail test of significance, significant. Table 3 summarizes those findings.

One may observe from Table 3 that mean number of words seems to be the feature most consistent with the holistic scoring. While the number of T-units seems to grow in importance as a predictor of quality in high school, the number of words per T-unit, that is the length of T-units, becomes non-significantly correlated with holistic scoring after the eighth grade. Of the indices of invention, none seems to

form any significant patterns except possibly the number of non-clothing items, which begins to correlate significantly with quality scoring in the final grades of high school. We may also note that no feature is significantly correlated in the ninth grade, marking it as an important anomoly associated with patterns of regression and plateauing described earlier.

Discussion

It appears that mere length of discourse (not length of T-units, however) is the best predictor of quality, at least in this study. Though both holistic and analytic instruments showed significant correlation overall, using the total sample of 125 papers, correlation of these factors by grade level revealed important irregularities, particularly with respect to regressions observed between grades seven and ten. Significant correlations by grade level also showed at what specific periods each of our quantitative features was in significant agreement with the quality scoring. Both these observations suggest that correlation studies may be needed to refine our understanding of the growth and improvement of writing and its assessment, not in terms of single variables, but rather as an interaction of several features that grow and recede as the child's written discourse evolves grade to grade.

This study has limitations. First, the general ninth grade population from which the sample was drawn, may have been different from the other grade level populations. This could explain the lack of correlation between quality and any of the six extrinsic variables.

However, when investigators undertake a regional writing assessment they are constrained by the teacher's response to the "call for papers." Second, it would have been better to have had papers from sixth graders to fill in, or better explain, the gap between fifth and seventh grade. Third, investigators may have been arbitrary in selecting the three variables of invention. Other variables might have produced different results. Despite these shortcomings researchers can learn a lot about student writing with cross-sectional analysis of student papers all on the same topic.

References

Cooper, C. R. & Odell, L. Evaluating Writing. Urbana: NCTE, 1977

Hunt, K. W. Grammatical structures written at three grade levels. Research Report No. 3. Urbana: NCTE, 1965.

Hunt, K. W. Syntactic maturity in school children and adults. Monographs of the Society for Research in Child Development, Serial No. 134, Vol. 35, No. 3, 1970.

Loban, W. D. Language ability in the middle grades of elementary school. USOE Cooperative Research Project SAE 7287, 1961.

Loban, W. D. Language development: Kindergarten through grade twelve. ResearchReport No. 18. Urbana: NCTE, 1976.

Myers, M. A procedure for writing assessment and holistic evaluation. Urbana: NCTE, 1980.

O'Hare, F. Sentence combining. Research Report No. 15. Urbana: NCTE, 1973.

O'Donnel, R. C., Griffin, W. J. & Norris, R. C. Syntax of kindergarten and elementary school children. Research Report No. 8. Urbana: NCTE, 1967.

Table 1

Distribution of Random Samples of Twenty-five Essays from Each of Seven Grade Levels

Grade Level	Population	Sample	Number of Teachers Represented
4-5	66	25	2
7	62	25	2
8	186	25	6
9	226	25	8
10	173	25	5
11	93	25	4
12	47	25	2
Totals	853	175	29

Table 2

Pearson Correlation Coefficients

On Quality Score with 6 Quantifiable Variables

on the Tctal Sample of Composition

	Quality Score	N	2-Tail Test of Significance
No. of Words	.7925	175	.001
No. of T-Units	.6533	175	.001
Words per T-Unit	.3462	175	.001
No. of items	.3348	175	.001
No. of Items Described	.3309	175	.001
No. of Non-Clothing Items	.4260	175	.001

Table 3

Significant ($p<.05$) Correlations by Grade Level after Pearson Correlation Coefficients

	4-5	7	8	9	10	11	12
Syntactic Features							
No. of Words	.67	.63	.73	NS	.52	.75	.75
No. of T-Units	NS	NS	NS	NS	.63	.53	.66
Words per T-Unit	.52	.25	-.26	NS	NS	NS	NS
No. of Items	.47	NS	NS	NS	NS	NS	NS
No. of Items Described	.63	NS	NS	NS	NS	.39	NS
No. of Non-Clothing Items	NS	NS	NS	NS	NS	.50	.52

Instructions: Read the following story. Follow the directions that come after it.

The Lost Suitcase

> Your name is Terry Lee. Three weeks ago you took a short bus ride to Santa Barbara to visit your cousin. On your return home, you realized that one of your suitcases was lost. Mr. Sam Jones, manager of the local bus station, told you to write a letter to the Lost and Found Division of the Braddock Bus Company. Jones tells you that Ms. Mary Martin, Director of the Lost and Found Division will help you find your suitcase. Her business address is 1451 West First Street, Los Angeles, California, 95321.

Directions for Writing

Write a letter to Ms. Martin, asking her to help you in finding your suitcase. You will need to (1) describe what the suitcase looked like, (2) describe the items that were inside the suitcaes, and (3) tell how you think the suitcase might have got lost. Although you will sign the letter as Terry Lee, you should put your own name in the top right-hand corner of your letter.

Time Limit: 40 minutes

Fig. 1. The topic used to assess the writing abilities.

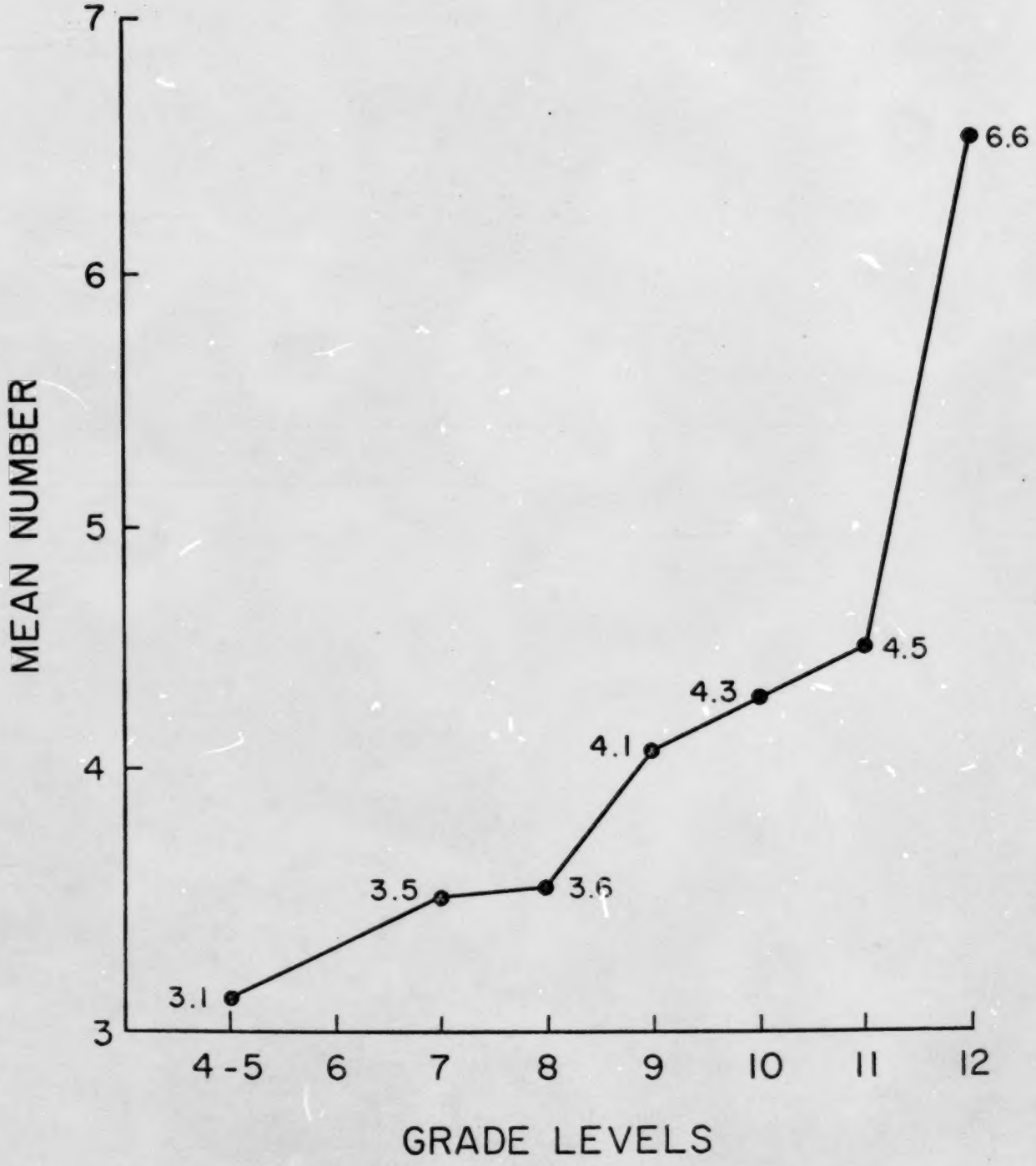

Figure 2: MEAN HOLISTIC SCORE ACROSS GRADE LEVELS

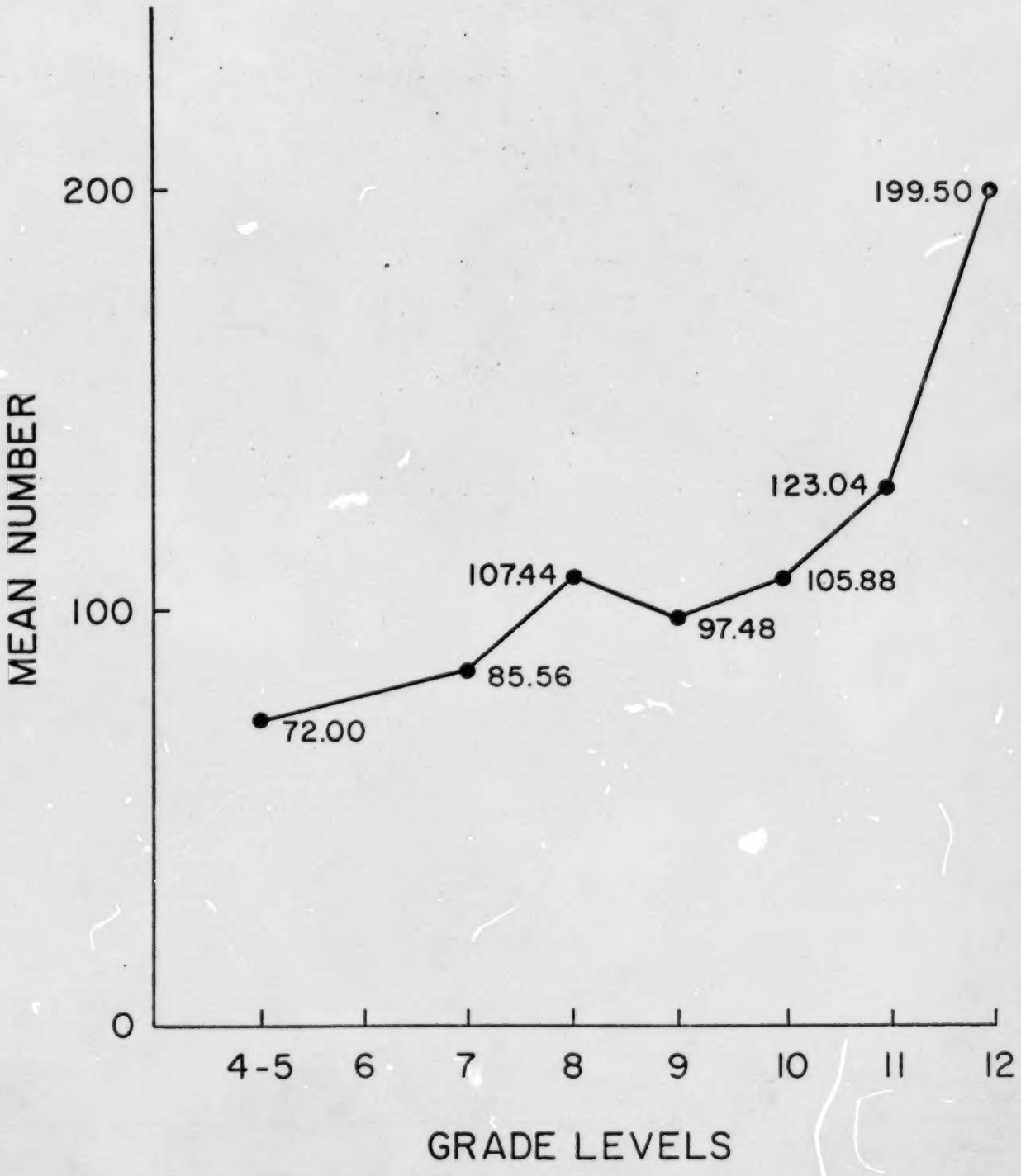

Figure 3: MEAN WORDS ACROSS GRADE LEVELS

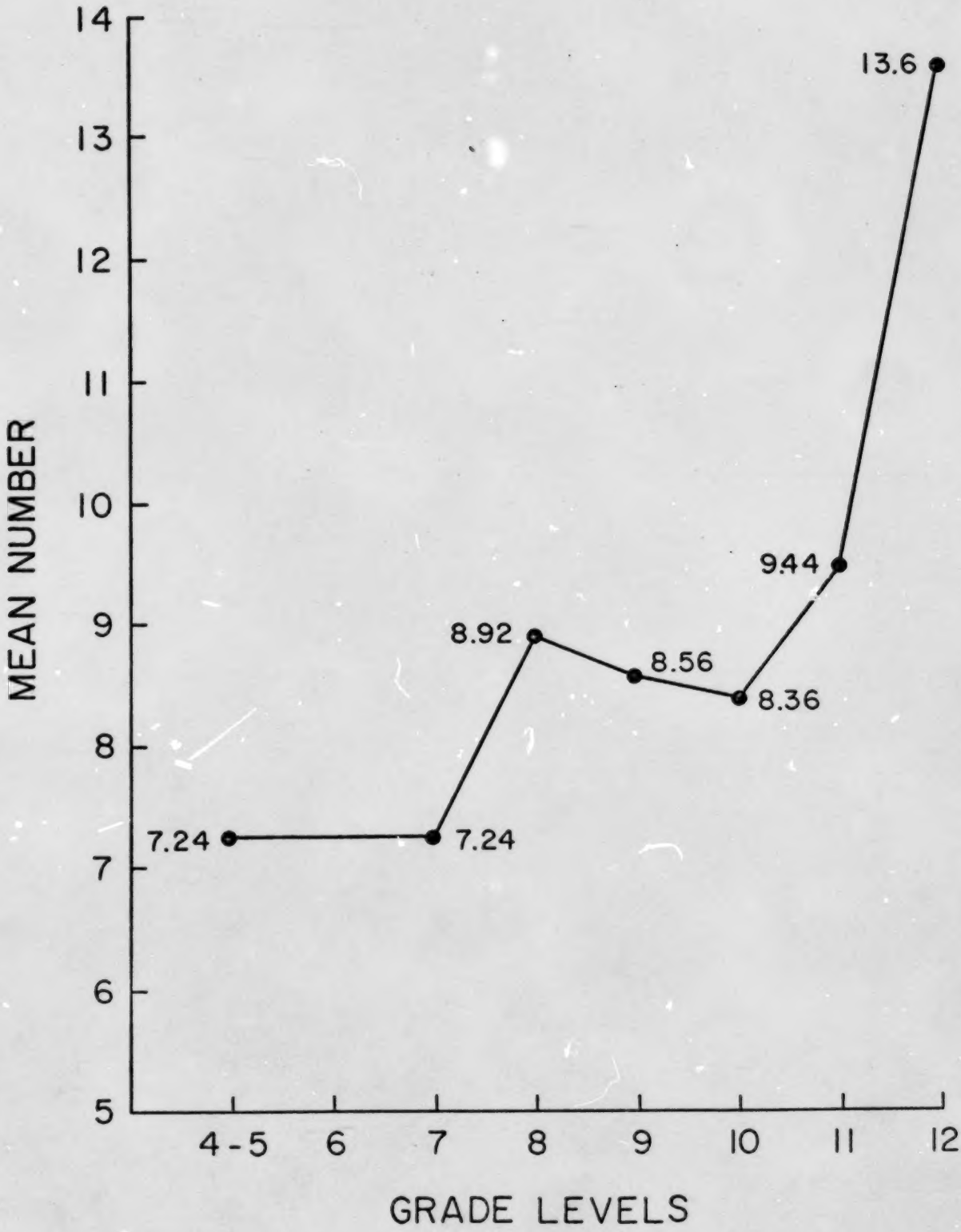

Figure 4. MEAN NUMBER OF T-UNITS ACROSS GRADE LEVELS

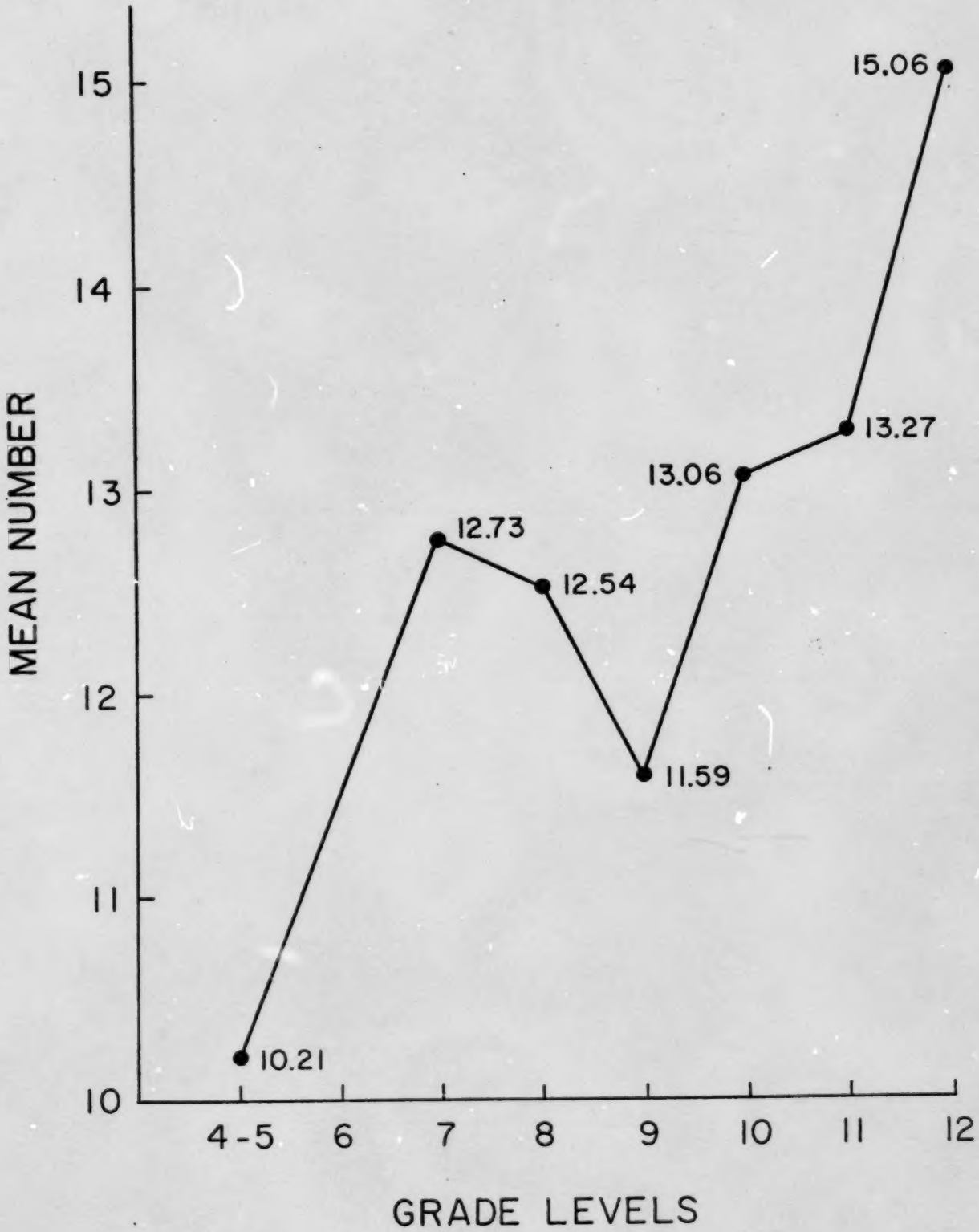

Figure 5: MEAN WORDS/T-UNITS ACROSS GRADE LEVELS

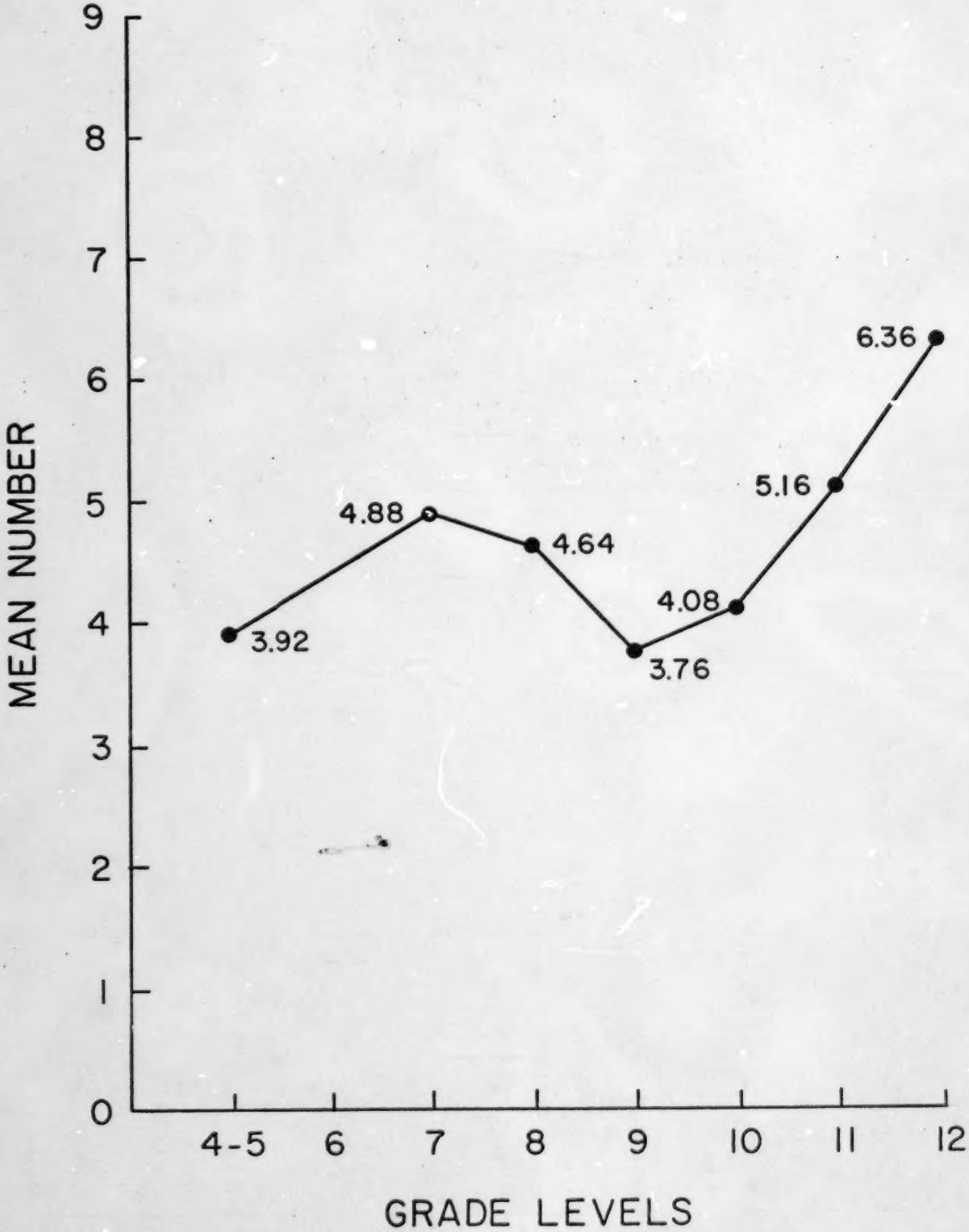

Figure 6: MEAN NUMBER OF ITEMS ACROSS GRADE LEVELS

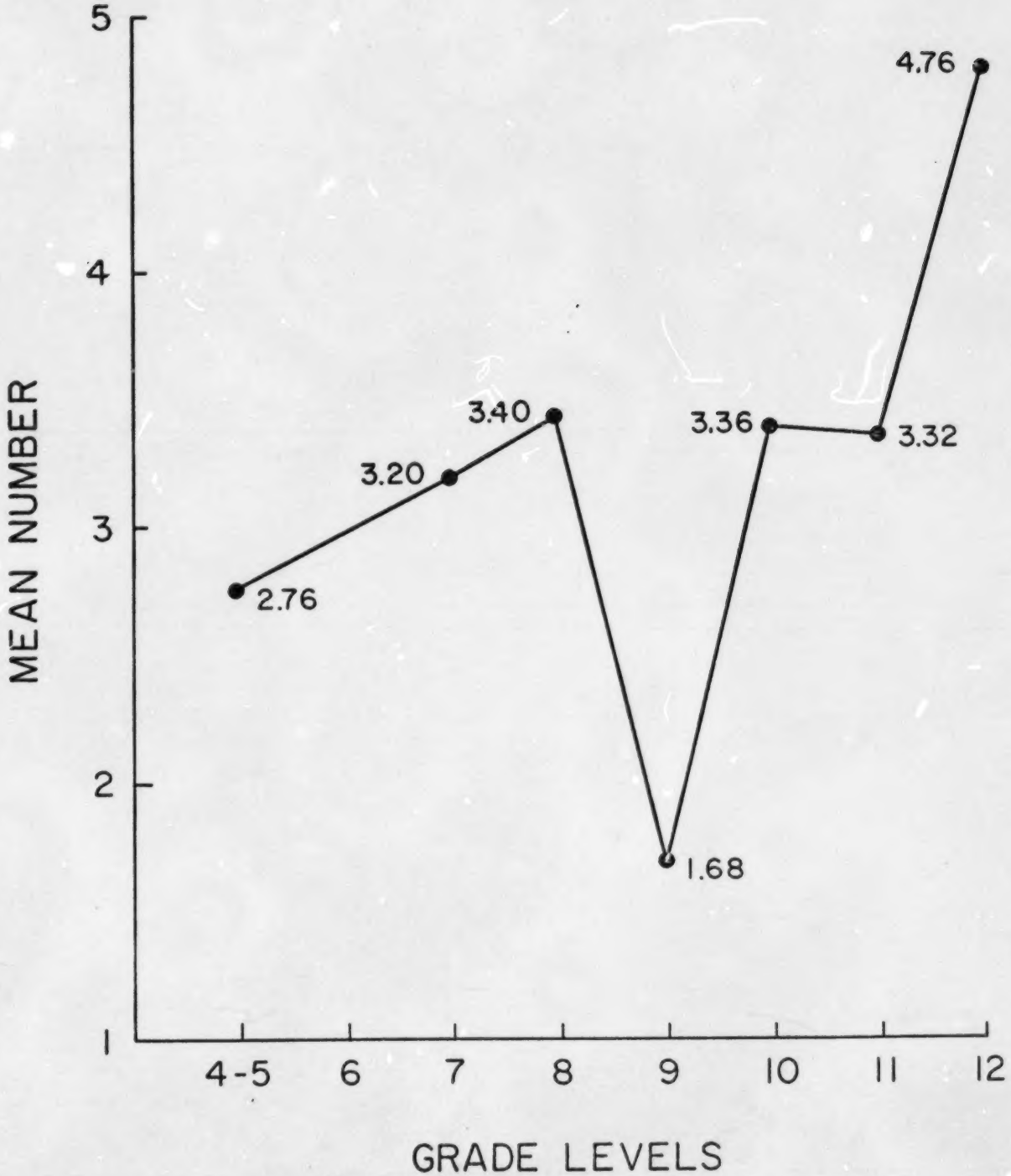

Figure 7: MEAN NUMBER OF ITEMS DESCRIBED ACROSS GRADE LEVELS

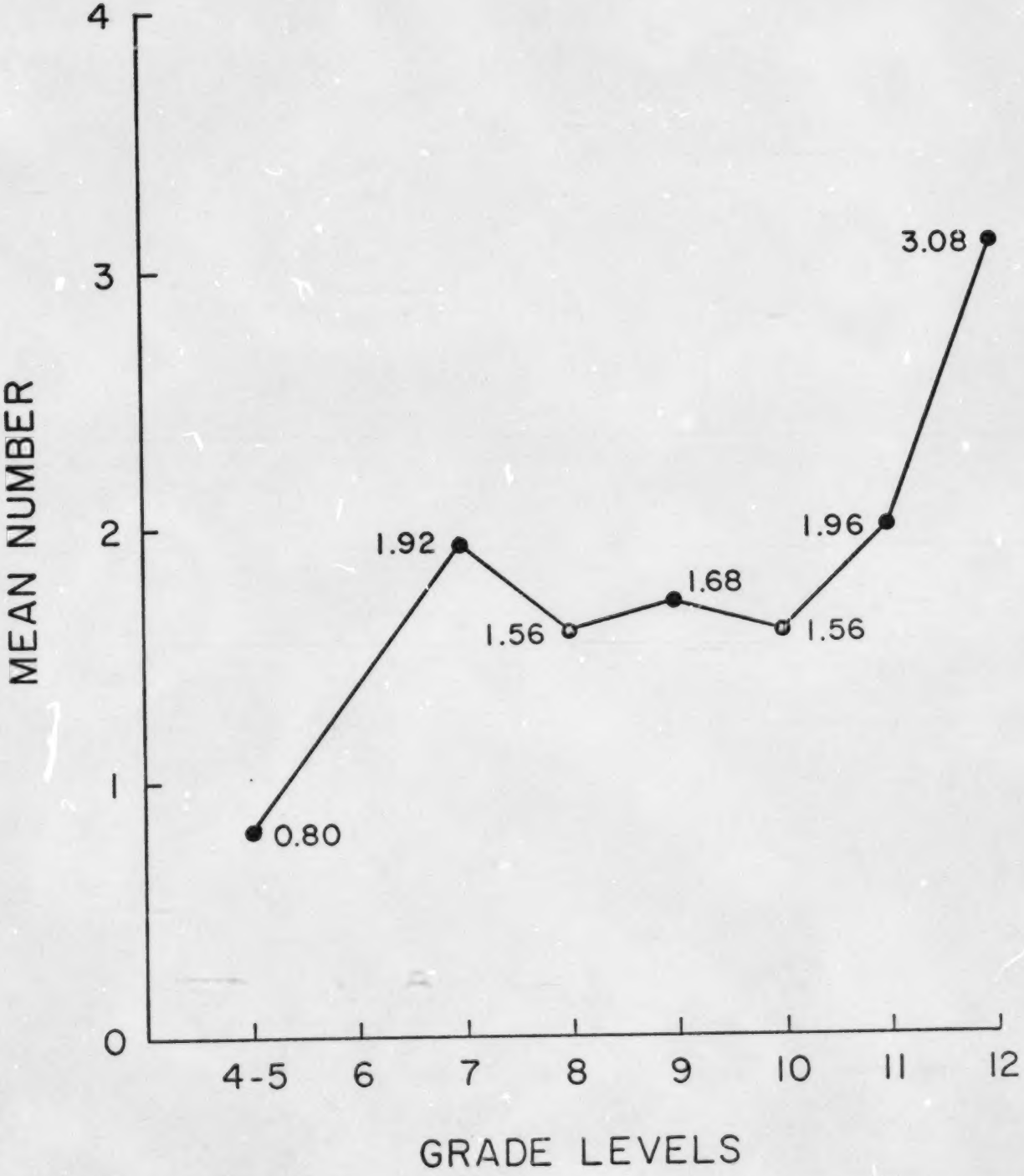

Figure 8 : MEAN NUMBER OF NON-CLOTHING ITEMS ACROSS GRADE LEVELS

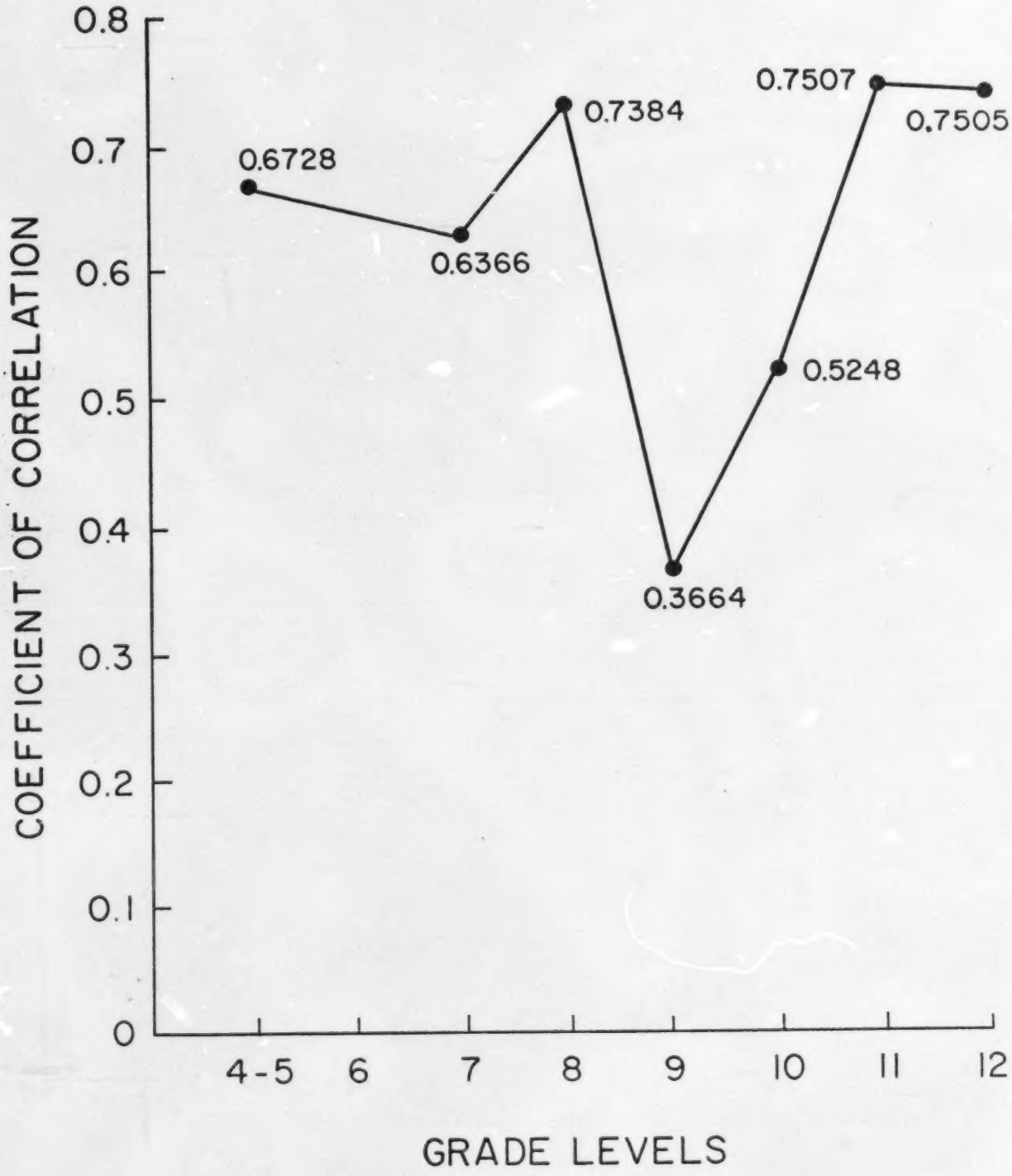

Figure 9 : CORRELATION OF HOLISTIC SCORES WITH MEAN NUMBER OF WORDS BY GRADE LEVELS

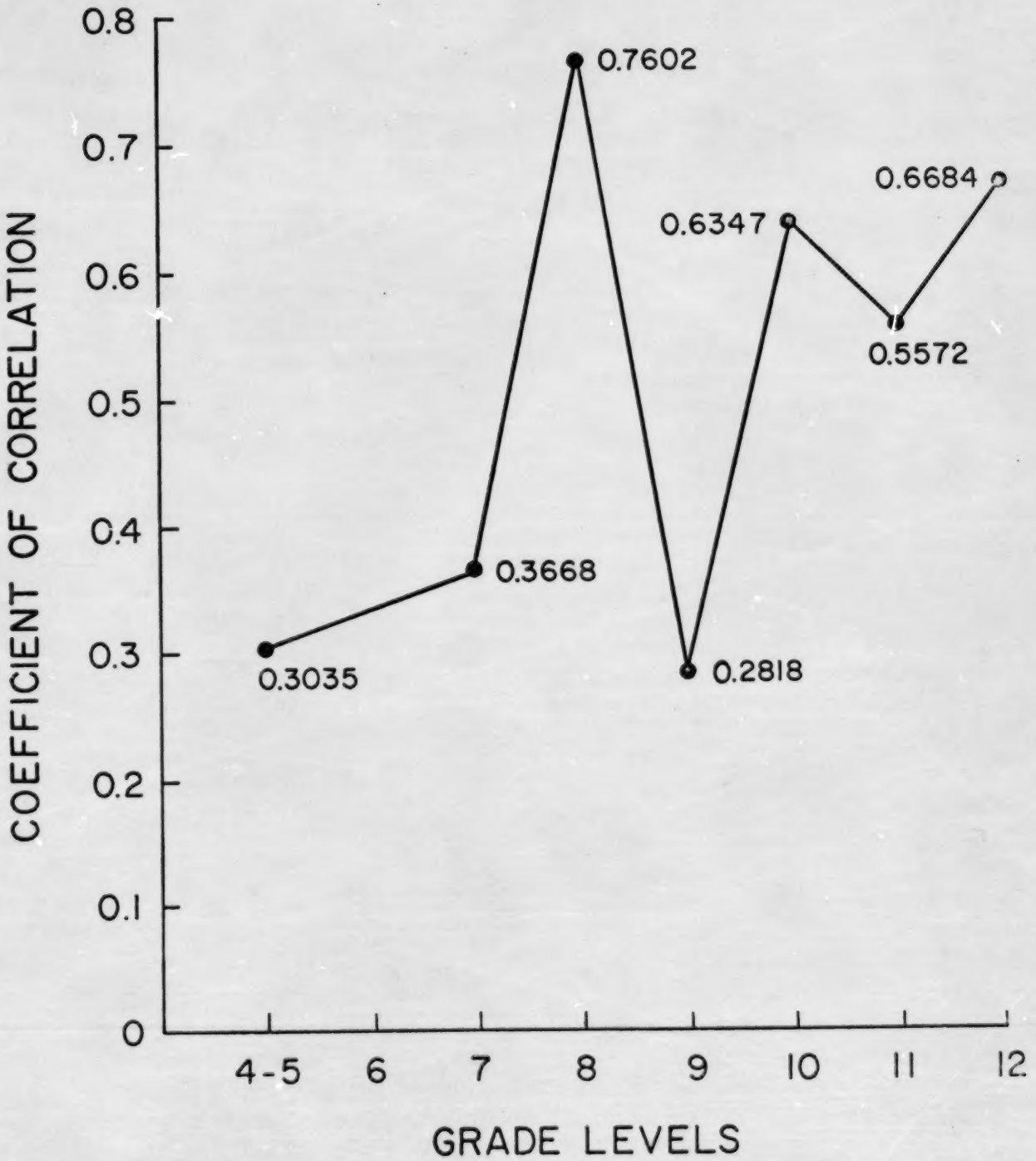

Figure 10: CORRELATION OF HOLISTIC SCORES WITH MEAN NUMBER OF T-UNITS BY GRADE LEVELS

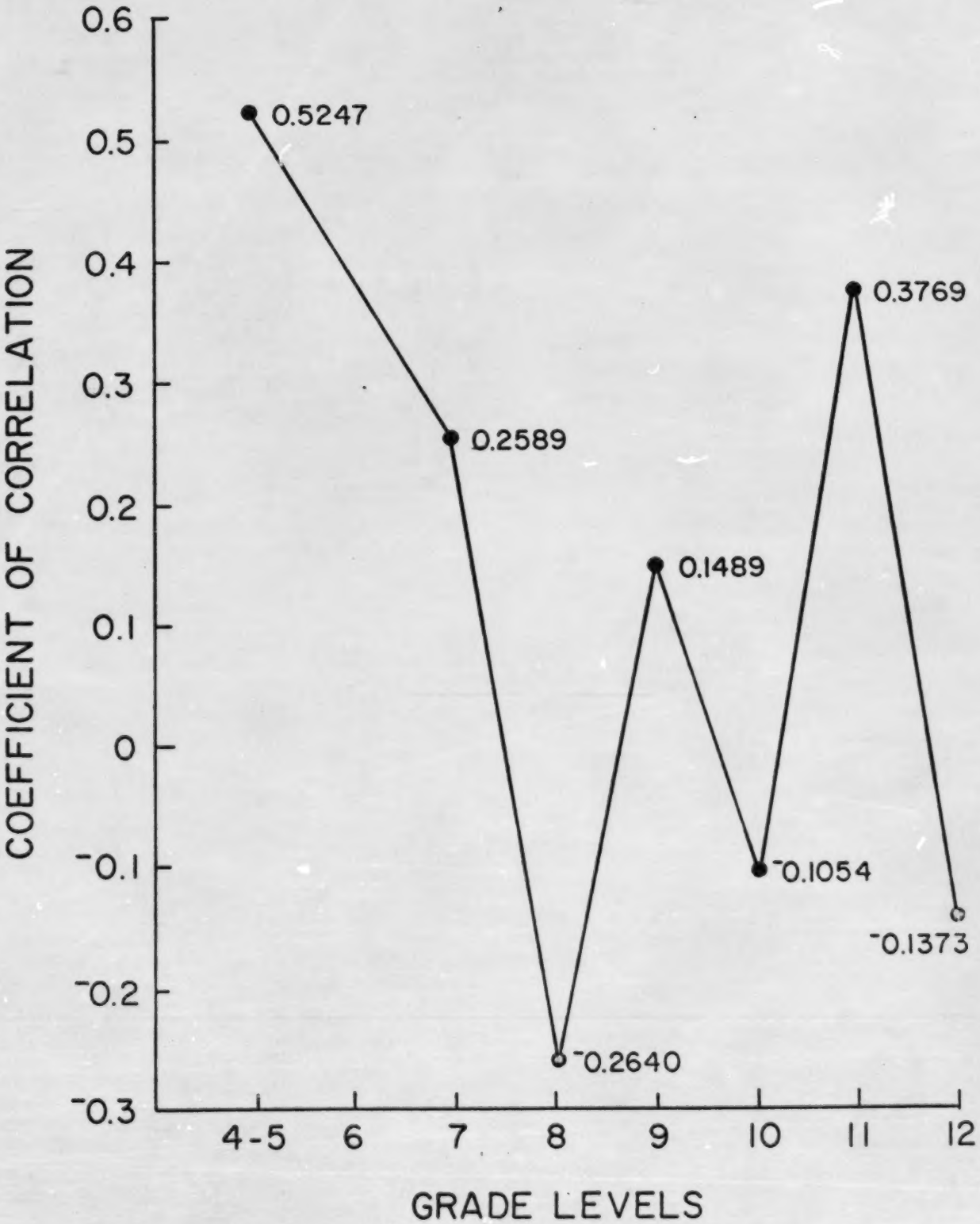

Figure 11: CORRELATION OF HOLISTIC SCORES WITH NUMBER OF WORDS/T-UNIT OVER GRADE LEVELS

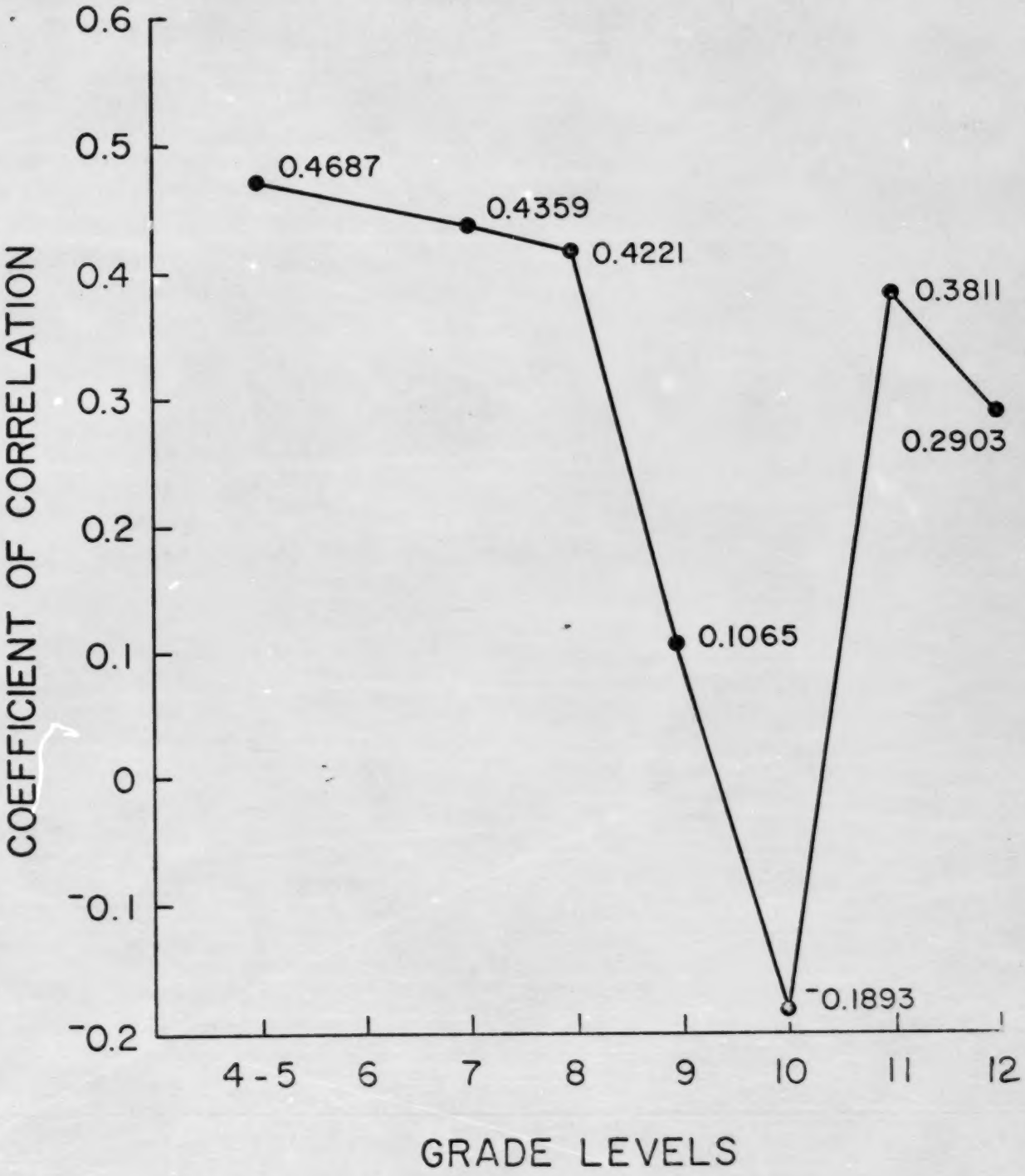

Figure 12: CORRELATION OF HOLISTIC SCORES WITH NUMBER OF ITEMS BY GRADE LEVELS

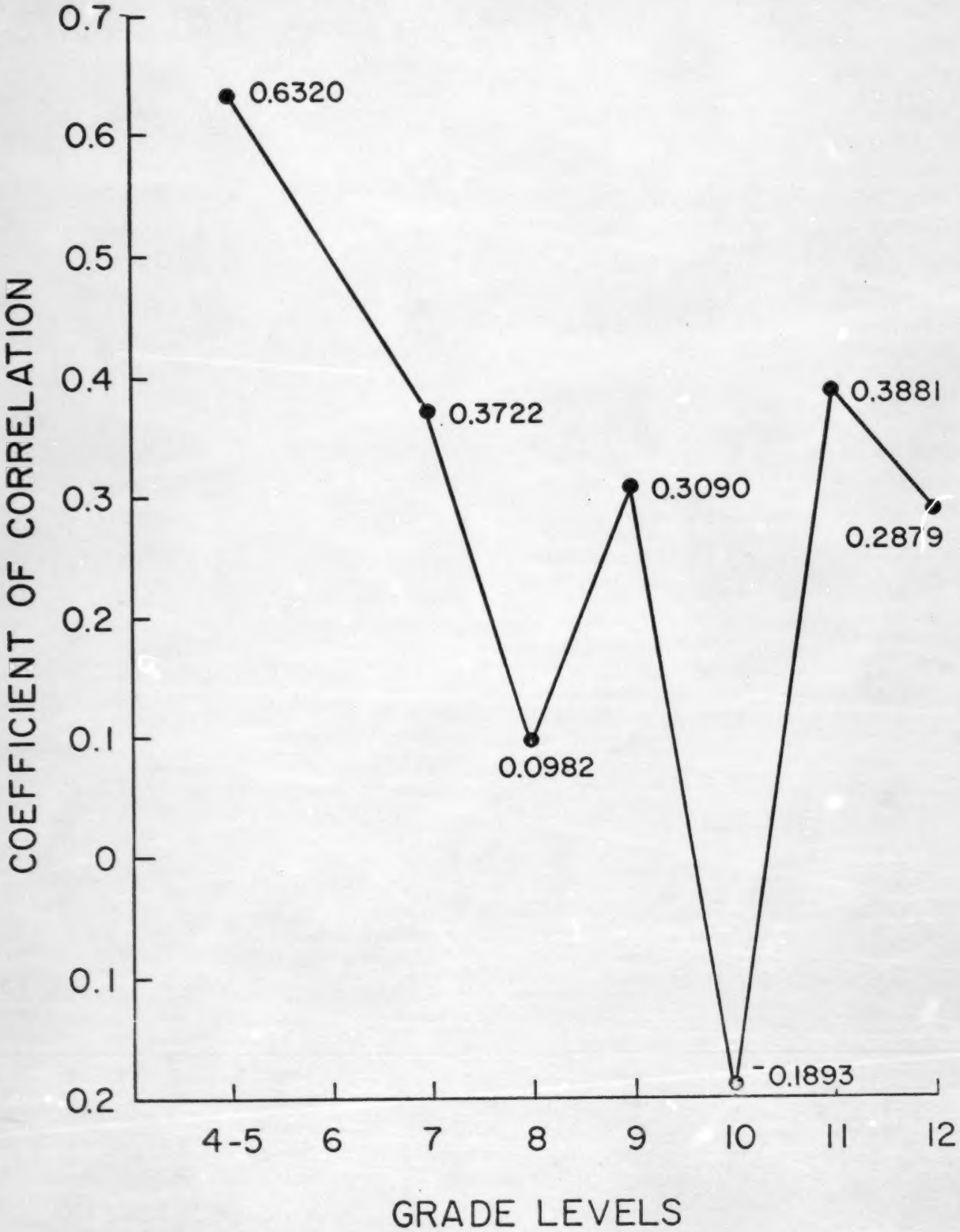

Figure 13: CORRELATION OF HOLISTIC SCORES WITH NUMBER OF ITEMS DESCRIBED BY GRADE LEVELS

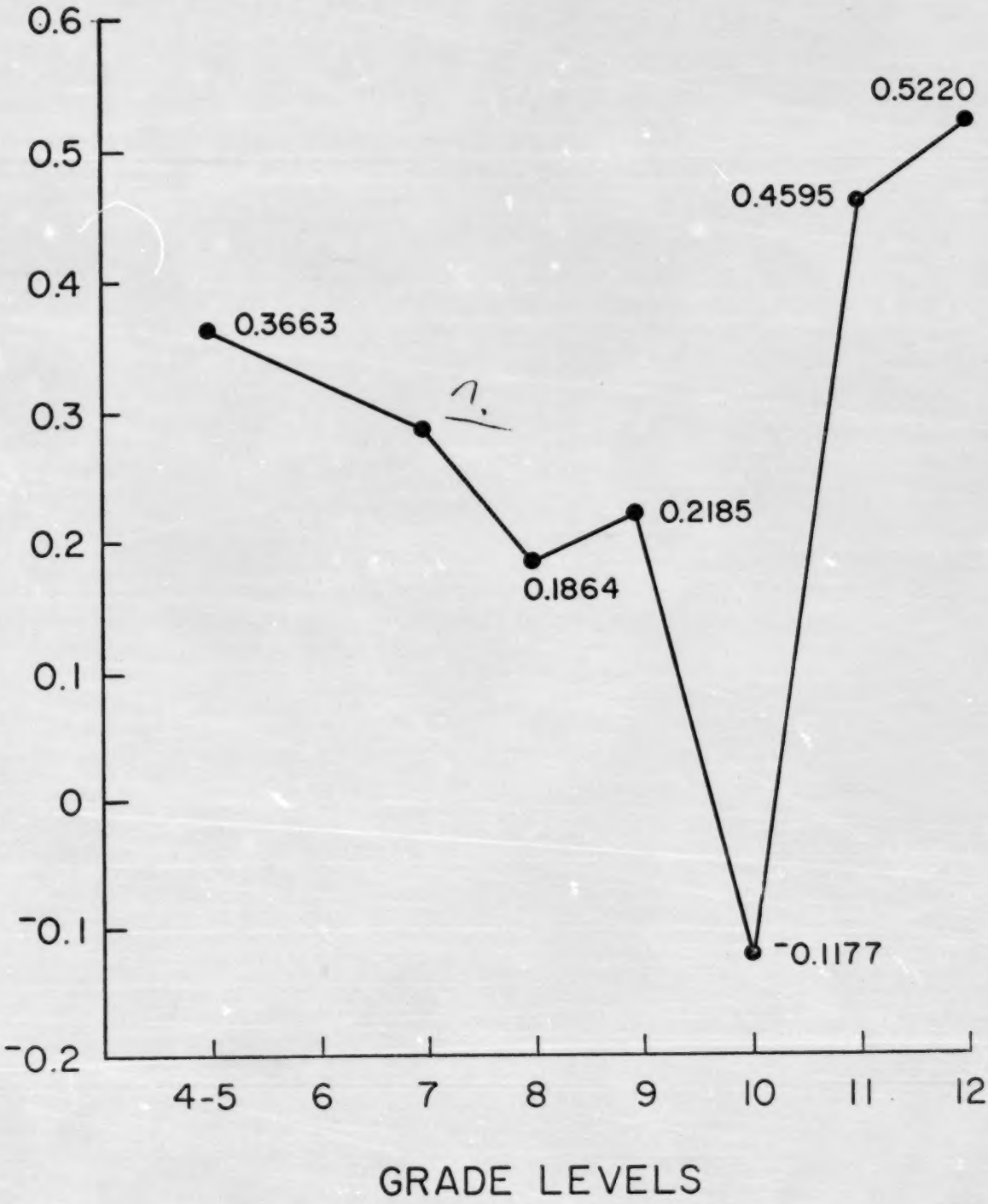

Figure 14: CORRELATION OF HOLISTIC SCORES WITH NUMBER OF ITEMS, NON-CLOTHING, BY GRADE LEVELS